

Amendments to the Claims:

1. (currently amended) A computer-implemented method for distributing data items from a particular set of data into a plurality of buckets based on distribution key[[s]] values associated with said data items, the method comprising the steps of: randomly selecting data items from said particular set of data to produce a sampled set of data items; determining a range for each bucket of said plurality of buckets based on a number of the distribution key[[s]] values associated with said sampled set of data items that fall within said range; assigning said range to each bucket of said plurality of buckets; and distributing each data item in said particular set of data to the bucket that has been assigned the range into which falls the distribution key value of the data item.

2. (currently amended) A computer-implemented method for distributing data items from a particular set of data into a plurality of buckets based on distribution key[[s]] values associated with said data items, the method comprising the steps of: randomly selecting data items from each subset of a plurality of subsets of said particular set of data to produce a sampled set of data items; determining a plurality of ranges based on the distribution key[[s]] values associated with the sampled set of data items; assigning said plurality of ranges to said plurality of buckets; and distributing each data item in said particular set of data to the bucket that has been assigned the range into which falls the distribution key value of the data item.

1 3. (original) The method of Claim 2 wherein the step of randomly selecting data items
2 from each subset of a plurality of subsets of said particular set of data includes
3 randomly selecting data items from each partition of a partitioned table.

1 4. (original) The method of Claim 2 wherein the step of randomly selecting data items
2 from each subset of a plurality of subsets of said particular set of data includes
3 randomly selecting data items from subsets of data, stored in buffers in volatile
4 memory, that represent results of one or more previously performed operations.

1 5. (original) The method of Claim 1 further comprising the steps of:
2 assigning the plurality of buckets to a plurality of processes; and
3 causing each process of said plurality of processes to perform, in parallel with the
4 other processes of said plurality of processes, an operation on the data items
5 contained in any buckets assigned to the process.

1 6. (original) The method of Claim 2 further comprising the step of selecting a distinct
2 random seed for each subset of the plurality of subsets of said particular set of data.

1 7. (currently amended) A computer-implemented method for distributing data items
2 from a particular set of data into a plurality of buckets based on distribution key[[s]]
3 values associated with said data items, the method comprising the steps of:
4 randomly selecting data items from said particular set of data to produce a sampled set
5 of data items; wherein
6 the particular set of data is durably stored on a plurality of durable
7 storage units; and

8 the step of randomly selecting data items includes randomly selecting
9 durable storage units from said plurality of durable storage units
10 and using the data items stored on said randomly selected
11 durable storage units as the sampled set of data items;
12 determining a plurality of ranges based on the distribution key[[s]] values associated
13 with the sampled set of data items;
14 assigning said plurality of ranges to said plurality of buckets; and
15 distributing each data item in said particular set of data to the bucket that has been
16 assigned the range into which falls the distribution key value of the data item.

1 8. (original) The method of Claim 1 wherein the step of randomly selecting data items
2 includes selecting a specified percentage of data items in said particular set of data.

1 9. (original) The method of Claim 7 wherein the step of randomly selecting data items
2 includes selecting a specified percentage of the plurality of durable storage units that
3 are storing said particular set of data.

1 10. (original) The method of Claim 8 further comprising the step of receiving, from a
2 user, data that specifies said percentage.

1 11. (original) The method of Claim 9 further comprising the step of receiving, from a
2 user, data that specifies said percentage.

1 12. (original) The method of Claim 5 wherein said operation is specified in a database
2 command, the method further comprising receiving with said database command data

3 that indicates how much of said particular set of data to randomly select to produce
4 said sampled set of data items.

1 13. (currently amended) The method of Claim 1 wherein said ranges contain an
2 approximately equal number of distribution key[[s]] values associated with said
3 sampled set of data items.

1 14. (currently amended) A computer-readable medium carrying instructions for
2 distributing data items from a particular set of data into a plurality of buckets based on
3 distribution key[[s]] values associated with said data items, the instructions
4 comprising instructions for performing the steps of:
5 randomly selecting data items from said particular set of data to produce a sampled set
6 of data items;
7 determining a range for each bucket of said plurality of buckets based on a number of
8 the distribution key[[s]] values associated with said sampled set of data items
9 that fall within said range;
10 assigning said range to each bucket of said plurality of buckets; and
11 distributing each data item in said particular set of data to the bucket that has been
12 assigned the range into which falls the distribution key value of the data item.

1 15. (currently amended) A computer-readable medium carrying instructions for
2 distributing data items from a particular set of data into a plurality of buckets based on
3 distribution key[[s]] values associated with said data items, the instructions
4 comprising instructions for performing the steps of:

5 randomly selecting data items from each subset of a plurality of subsets of said
6 particular set of data to produce a sampled set of data items;
7 determining a plurality of ranges based on the distribution key[[s]] values associated
8 with the sampled set of data items;
9 assigning said plurality of ranges to said plurality of buckets; and
10 distributing each data item in said particular set of data to the bucket that has been
11 assigned the range into which falls the distribution key value of the data item.

1 16. (original) The computer-readable medium of Claim 15 wherein the step of randomly
2 selecting data items from each subset of a plurality of subsets of said particular set of
3 data includes randomly selecting data items from each partition of a partitioned table.

1 17. (original) The computer-readable medium of Claim 15 wherein the step of randomly
2 selecting data items from each subset of a plurality of subsets of said particular set of
3 data includes randomly selecting data items from subsets of data, stored in buffers in
4 volatile memory, that represent results of one or more previously performed
5 operations.

1 18. (original) The computer-readable medium of Claim 14 further comprising instructions
2 for performing the steps of:
3 assigning the plurality of buckets to a plurality of processes; and
4 causing each process of said plurality of processes to perform, in parallel with the
5 other processes of said plurality of processes, an operation on the data items
6 contained in any buckets assigned to the process.

1 19. (original) The computer-readable medium of Claim 15 further comprising instructions
2 for performing the step of selecting a distinct random seed for each subset of the
3 plurality of subsets of said particular set of data.

1 20. (currently amended) A computer-readable medium carrying instructions for
2 distributing data items from a particular set of data into a plurality of buckets based on
3 distribution key[[s]] values associated with said data items, the instructions
4 comprising instructions for performing the steps of:

5 randomly selecting data items from said particular set of data to produce a sampled set
6 of data items; wherein

7 the particular set of data is durably stored on a plurality of durable
8 storage units; and

9 the step of randomly selecting data items from said particular set of
10 data to produce a sampled set of data items includes randomly
11 selecting durable storage units from said plurality of durable
12 storage units and using the data items stored on said randomly
13 selected durable storage units as the sampled set of data items;

14 determining a plurality of ranges based on the distribution key[[s]] values associated
15 with the sampled set of data items;

16 assigning said plurality of ranges to said plurality of buckets; and

17 distributing each data item in said particular set of data to the bucket that has been

18 assigned the range into which falls the distribution key value of the data item.

- 1 21. (original) The computer-readable medium of Claim 14 wherein the step of randomly
2 selecting data items includes selecting a specified percentage of data items in said
3 particular set of data.
- 1 22. (original) The computer-readable medium of Claim 20 wherein the step of randomly
2 selecting data items includes selecting a specified percentage of the plurality of
3 durable storage units that are storing said particular set of data.
- 1 23. (original) The computer-readable medium of Claim 21 further comprising instructions
2 for performing the step of receiving, from a user, data that specifies said percentage.
- 1 24. (original) The computer-readable medium of Claim 22 further comprising instructions
2 for performing the step of receiving, from a user, data that specifies said percentage.
- 1 25. (original) The computer-readable medium of Claim 18 wherein said operation is
2 specified in a database command, the computer-readable medium further comprising
3 instructions for receiving with said database command data that indicates how much
4 of said particular set of data to randomly select to produce said sampled set of data
5 items.
- 1 26. (currently amended) The computer-readable medium of Claim 14 wherein said ranges
2 contain an approximately equal number of distribution key[[s]] values associated with
3 said sampled set of data items.